

СТВОРЕННЯ ТА ВИКОРИСТАННЯ ІНТЕГРАЛЬНИХ ПРОГНОЗНИХ МОДЕЛЕЙ ДЛЯ ПІДВИЩЕННЯ АДЕКВАТНОСТІ РЕЗУЛЬТАТІВ АВТОМАТИЗОВАНОГО ГЕОЛОГІЧНОГО ПРОГНОЗУВАННЯ

Наведено теоретичні основи комп'ютерної технології прогнозування локалізації геологічних об'єктів за комплексом геолого-геофізичних і космічних даних. Основу технології складають процеси формування інтегральних прогнозних моделей, що поєднують матеріали наземних і дистанційних зйомок, і їх застосування для одержання прогнозів з використанням методів Data Mining.

Приведены теоретические основы компьютерной технологии прогнозирования локализации геологических объектов по комплексу геолого-геофизических и космических данных. Основу технологии составляют процессы формирования интегральных прогнозных моделей, объединяющих материалы наземных и дистанционных съемок, и их применения для получения прогнозов с использованием методов Data Mining.

The theoretical basics of computer technology for prediction of geological objects localization from a set of geological, geophysical and satellite data is presented. The technology is based processes of integrated predictive models forming that combine data of ground-based observations and remote surveys, and their application for prediction tasks using Data Mining methods.

Стан проблеми. Теорія та практика автоматизованого прогнозування локалізації геологічних об'єктів, що пережили фазу бурхливого розвитку в 70–90 роках пришлого століття, в наш час перебувають у стані, який можна визначити як стагнація. Постійний ріст об'ємів геоданих, що добуваються, у першу чергу, космічних, не супроводжується настільки ж швидким розвитком методів спільної обробки й аналізу величезних масивів інформації. Як наслідок, фахівці-геологи зіштовхуються з низкою технологічних, методологічних і теоретичних труднощів, що обумовлює підвищення складності й зниження ефективності сумісної обробки й аналізу величезних об'ємів різнорідних і різнорівневих даних. У результаті практичне використання потенційно корисних матеріалів найчастіше обмежується їх якісним візуальним аналізом.

До теперішнього часу відсутня єдина точка зору на ключові методологічні питання комп'ютерного прогнозування – вибору підходів до побудови вирішальних правил і інформативних систем ознак, оцінці інформативності і якості прогнозів і ін. Існує явний дисбаланс між рівнями розвитку математичного й методологічного апаратів комп'ютерного прогнозування в науках про Землю. Кількість методів і алгоритмів, що вирішують окремі задачі, обчислюється тисячами, але створених на їхній базі цільних, завершених методик, доведених до стадії технології, надзвичайно мало. При цьому більшість із технологічних схем, що рекомендуються, складаються з етапів, які мають настільки загальний і абстрактних характер, що застосування їх на практиці вкрай утруднено.

Практично відсутні сучасні спеціалізовані технології і системи автоматизованого прогнозування. Як наслідок, виробничі організації найчастіше використовують універсальний інструментарій прогнозування, реалізований у таких

системах, як ERDAS, ArcGIS і т.п. і слабо адаптований до рішення реальних прогнозних геологічних задач.

Крім того, практичне застосування існуючої теоретичної бази автоматичної прогнозування локалізації об'єктів на засадах апарату класифікації стримується низкою принципових, специфічних для геологічної галузі труднощів, серед яких: малі обсяги еталонних вибірок; невисокий ступінь подібності об'єктів, що вважаються однотипними; наявність величезної кількості перешкод у даних дистанційних методів та їх досить низька інформативність та інші.

Мета роботи. Для подолання зазначених труднощів запропонована спеціалізована геоінформаційна технологія [1] рішення прогнозно-пошукових геологічних завдань, що базується на інтеграції, з одного боку, різнорідних і різновисотних геоданих – матеріалів майданних геофізичних, геохімічних, аерокосмічних, топографічних зйомок, картографічних джерел, цифрових геологічних моделей і ін., та, з іншого боку, моделей та методів геоінформатики та Data Mining (інакше – інтелектуального аналізу даних) [2, 3]. Процес прогнозування в рамках технології зводиться до побудови т.зв. інтегральних прогнозних моделей (ІПМ) та отримання на їх основі прогнозів методами Data Mining.

Для побудови ІПМ розроблено низку методик та практичних прийомів, які мають на меті підвищення адекватності прогнозів, що отримуються. Робота присвячена викладенню теоретичних основ побудови та використання ІПМ [4].

Основні відомості про ІПМ

В основі технології автоматизованого прогнозування лежать:

- інтегральна прогнозна модель (ІПМ) на основі описових і прогнозних моделей Data Mining;
- методика формування ІПМ, що має на меті максимальне узгодження її складових елементів для одержання геологічно змістовних прогнозів;
- методи інтелектуального аналізу різнорідних та різнорівневих геоданих для одержання нових знань на основі ІПМ.

Інтегральна прогнозна модель поєднує:

- математичні, геоіконічні та геологічні моделі;
- матеріали різнорідних та різнорівневих космічних та наземних зйомок;
- описові і прогнозні моделі Data Mining;
- різноманітні процедури обробки й аналізу даних.

При формуванні ІПМ переслідується ціль мінімізації неузгодженості геоданих і геологічних фактів. Під **неузгодженістю** розуміється нездатність даних непрямих і дистанційних методів вивчення Землі адекватно описувати геологічне середовище. Вона є наслідком недостатньої інформативності, глибинності й роздільної здатності окремих геофізичних і аерокосмічних методів, адитивності потенційних геофізичних полів, наявності різноманітних шумових компонентів негеологічного походження й т.п. На практиці неузгодженість веде до внутрішньої суперечливості прогнозних моделей і відчутно знижує надійність одержуваних рішень. Вплив неузгодженості на практиці звичайно неможливо усунути повністю, але можна зменшити із застосуванням спеціальних прийомів. Мінімізація неузгодженості дозволяє суттєво підвищити вірогідність й змістов-

ність результатів рішення прогнозно-пошукових задач, зокрема, автоматизованого прогнозування локалізації геологічних об'єктів та явищ.

ІПМ являє собою сукупність взаємодіючих описових (*ОМ1*, *ОМ2* и *ОМ3*) й прогнозної моделі (*ІПМ*), які складаються з базових модельних елементів (рис. 1), серед яких:

- такі, що формалізують геологічні факти й гіпотези:
 - моделі природно-геологічного поділу (*ІПП*);
 - цифрові геологічні модель 1-го (*ЦГМ1*) та 2-го (*ЦГМ2*) типів;
- геодані: двовимірні характеристики геологічного середовища, що формують багатомірний ознаковий простір (*ОП*);
- частини *ІПМ*, що підлягають оптимізації:
 - навчальні (*НВ*) та контрольні (*КВ*) вибірки;
 - поділяючі (*ПН*) та діагностуючі (*ДН*) набори ознак;
 - вирішальні правила (*ВП*);
- прогнози: двовимірні карти, що відбивають перспективність територій на виявлення певних геологічних об'єктів;
- критерії оптимальності і якості:
 - критерії геологічної змістовності (*ІГЗ*) і формальної інформативності (*ІФІ*) окремих ознак; критерії геологічної змістовності (*ГГЗ*), а також формальної інформативності (*ФІ*) систем (груп, наборів) ознак; критерії геологічної змістовності (*ГЗ*), формальної точності (*ФТ*) і надійності прогнозів.

Такі елементи прогнозних моделей як *ОВ*, *НВ*, *ДН*, *ВП*, *ОП*, *ФІ*, *ІФІ* та *ФТ* є типовими для прогнозуючих моделей та систем, то використання *ІПП*, *ЦГМ1*, *ЦГМ2*, *ІГЗ*, *ПН* є характерною відмінністю *ІПМ*. Нижче дана стисла характеристика основних елементів.

ІПП формалізує знання про відомі об'єкти прогнозованого типу (наприклад, вуглеводневі поклади, золоторудні об'єкти) і задає їхній поділ на класи (типи) відповідно до деякої класифікації, що відбиває об'єктивно існуючі й істотні для рішення поставленої задачі риси об'єктів (наприклад, поділ вуглеводневих покладів на нафтові, газові, газоконденсатні; або на пов'язані зі структурними й неструктурними пастками й т.п.). Модель *ІПП* є сукупністю об'єктів (осередків мережі), розділених експертом-геологом на *N* класів K_1, K_2, \dots, K_N відповідно до геологічних фактів або гіпотез. Також модель *ІПП* містить клас *K'* «порожніх» об'єктів, що завідомо не є цільовими.

ЦГМ1 і *ЦГМ2*, створювані фахівцем-геологом, формалізують площинну інформацію про геологічне середовище. Різниця між *ЦГМ1* і *ЦГМ2* – у повноті й масштабності інформації, що містяться в них. Якщо *ЦГМ2* узагальнює всі наявні достовірні геологічні знання про геологічне середовище, то *ЦГМ1* є похідною дрібномасштабних геологічних карт і містить відомості про геологічні структури низьких рангів. Масштаб карт, покладених в основу *ЦГМ1*, вибирається таким чином, щоб територія була розділена на 3–10 частин, що відповідають різним геологічним утворенням (літологічним різницям, стратиграфічним одиницям і т.п.). При необхідності, *ЦГМ1* може бути побудована шляхом генералізації *ЦГМ2*

ОП є сукупністю **ознак** – геоданих, організованих у сіткові шари. Значення ознак задані в комірках регулярної мережі, що покриває досліджувану територію. Розрізняються вхідні ознаки, отримані в результаті первинної обробки натурних зйомок (фізичні поля, геохімічні ореоли, КЗ й т.п.) і їхні трансформанти, отримані шляхом математичних перетворень вхідних ознак. Окремі класи становлять ознаки-характеристики геологічної будови, лінеаментних мереж і границь яскравості **геозображень** – геоданих, представлених графічно в растровій або сітковій формах.

Еталонні вибірки, що поділяються на **НВ** та **КВ**, формалізують інформацію про розташування відомих об'єктів розшукуваного типу (далі – **цільових об'єктів**) й формуються шляхом апроксимації комірками регулярної мережі проєкцій їхніх контурів на горизонтальну площину. Додатково використовуються еталонні вибірки, що містять об'єкти, які свідомо не належать до цільових (т.зв. нецільові або «порожні»). Структура й состав НВ і КВ визначаються в процесі формуванні ПМ.

Критерії індивідуальної і групової ГЗ служать для оцінки здатності ознак та їхніх наборів відбивати об'єктивно існуючу геологічну будову досліджуваної ділянки. ГЗ кількісно характеризує здатність групи ознак відбивати найбільш загальні риси геологічної будови, формалізовані в ЦГМ1.

ГЗ є певним чином розрахована міра подібності ознаки з ЦГМ2. Шляхом застосування процедури растеризації ЦГМ2 переводиться у форму бінарної карти, де одиницям відповідають стратиграфічні й літологічні границі, контури інтрузивних утворень, лінійні розривні порушення й т.п. Оцінювана ознака шляхом застосування метода Кенні [5] для виділення границь яскравості також переводиться в бінарну форму. Далі аналізується конфігурація малюнка кожної з карт із використанням ковзної околиці розміром 3x3 комірки. Для кожної карти будується гістограма частоти зустрічальності кожного з 255 можливих варіантів сполучення 0 і 1, після чого гістограми порівнюються у будь-який відомий спосіб (обчислення χ -квадрата, площі перетинання гістограм, відстані Бхаттачарія, кореляційним методом й т.п.). Отримані значення є оцінками міри подібності ознаки й ЦГМ2.

Критерії ФІ використовуються для оцінки здатності ознаки або групи ознак забезпечувати формальне рішення задачі класифікації з мінімальними втратами. ФІ оцінюється шляхом обчислення оцінки апіорної байєсовської ймовірності помилки класифікації (**ЙПК**) або подібних їй оцінок, в залежності від методу побудови ВП.

В якості **критеріїв ФТ** використовується група апостеріорних показників, заснованих на оцінці відповідності прогнозів наявним фактичним даним про цільові об'єкти. **ГЗ прогнозу** обчислюється подібно ГЗ ознак.

Поділяючий набір (ПН) – інформативна сукупність ознак, що забезпечують поділ об'єктів НВ у багатомірному просторі, оптимальний з погляду його близькості до ПП. Поділяючий набір будується при формуванні ОМ1. Його використання сприяє одержанню прогнозів, максимально погоджених з фактичною геологічною інформацією про цільові об'єкти.

Діагностуючий набір (ДН) – сукупність ознак, для якої досягається оптимальне з погляду ФІ й ГЗ рішення задачі класифікації з навчанням.

Формування ДН та НВ, за своєю суттю, зводиться до пошуку зв'язків між відомими цільовими об'єктами і геоданими, що їх описують. На основі виявлених зв'язків і сформульованих закономірностей може бути виконаний прогноз локалізації раніше невідомих об'єктів. Технологія припускає можливість виконання прогнозу на основі закономірностей, знайдених і іншими методами Data Mining – секвенційного аналізу, пошуку асоціації [2, 3] й т.п., однак метод класифікації є найбільш універсальним.

Знайдені закономірності формалізуються у вигляді вирішальних правил, що дозволяють оцінити приналежність об'єктів класифікації до одного або декількох заздалегідь визначених класів або визначити ступінь приналежності об'єкта до певного класу.

Перераховані базові елементи й зв'язки між ними формують більші частини ІПМ – ОМ1, ОМ2, ОМ3 і ПМ (рис. 1).

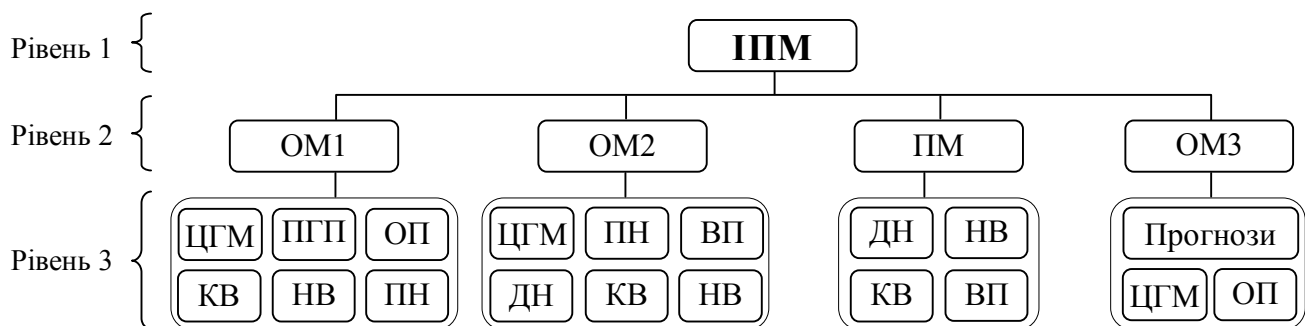


Рис. 1. Структура ІПМ

Методика формування ІПМ. Формування ІПМ зводиться до послідовного формування моделей ОМ1, ОМ2, ПМ та ОМ3.

Основна мета **побудови ОМ1** – забезпечення відповідності геоданих і фактичної геологічної інформації про відомі цільові об'єкти. Також за результатами побудови ОМ досягаються і інші цілі:

- створення ознакового простору, що забезпечує геологічно змістовний поділ НВ на підвибірки K_1, K_2, \dots, K_N ;
- кількісна оцінка придатності наявного ОП для рішення задач класифікації з необхідною погрішністю;
- формування контрольної вибірки й ряд інших задач.

Ключовим етапом створення ОМ1 є побудова поділяючого набору ознак (ПН). Для цього серед ОП (сукупності вхідних геоданих та їх трансформант) знаходиться набір ознак, який забезпечує кластеризацію об'єктів, максимально близьку до ПГП. Процедура виконується будь-яким з відомих методів пошуку оптимальних систем ознак (послідовного приєднання, послідовного відкидання, випадкового пошуку, випадковий пошук з адаптацією, приєднання з відкиданням, метод гілок і границь [6] і т.п.). Для цього організується ітераційний процес, і на кожному його кроці шукається набір, що мінімізує запропонований

показник ГЗ групи ознак: $QI = \frac{D_{BK}}{D_{MK}}$; $D_{BK} = \frac{\sum_i^n \sum_j^n d_E(x_i, x_j)}{2M}$; $D_{MK} = \frac{\sum_i^n \sum_j^n d_{mean}}{2N}$, де D_{BK}

– середня внутрішньокластерна відстань для N класів моделі ПГП; D_{MK} – середня міжкластерна відстань; x – об'єкт вибірки; M – кількість об'єктів; d_E – евклідова відстань; d_{mean} – середнє значення попарних відстаней між об'єктами різних класів. Розрахунок D_{BK} і D_{MK} виконується для об'єктів моделі ПГП у наявному ОП, за умови, що приналежність до кластерів задана ПГП.

Слід зазначити, що підхід з обчисленням QI дозволяє оцінювати можливість поділу об'єктів на кластери без проведення кластеризації. Це вимагає значно менших обчислювальних витрат.

Далі виконується кластеризація в просторі знайденого ПН. Кількість кластерів має бути рівною або близькою до кількості класів ПГП. Оцінюється розходження між отриманою кластеризацією й ПГП, яке характеризує придатність наявного ОП для досягнення класифікації, що мінімально суперечить геологічним фактам щодо цільових об'єктів. Якщо отримане значення розходження не задовольняє, розраховуються додаткові ознаки і процес повторюється.

За результатами кластеризації або багатомірного шкалювання в знайденому ОП виконується: а) перерозподіл об'єктів між вибірками K_1, K_2, \dots, K_N з метою підвищення компактності останніх; б) видалення з НВ об'єктів, віддалених від центрів всіх кластерів або таких, що попадають у різні кластери при використанні різних методів поділу. Вилучені об'єкти переводяться в КВ. У такий спосіб досягаються дві мети: 1) зниження апріорної ймовірності помилкової класифікації (ЙПК); 2) забезпечення контролю ФТ прогнозів.

Якщо на етапі формування ОМ1 упор робиться на вивчення цільових об'єктів, то при створенні ОМ2 – на їхні взаємини із вміщуючим середовищем. Побудова ОМ2 дозволяє відповісти на запитання – чи можливо зробити відділення об'єктів навчальних вибірок кожного з класів від «порожнього» вміщуючого середовища?

Крім того, в процесі створення ОМ2 досягаються наступні проміжні цілі: 1) оптимізація складу і структури НВ класів; 2) кількісна оцінка здатності ОП відбивати загальні риси геологічної будови території; 3) розширення КВ.

Побудова ОМ2. Якщо ставиться задача одержання категорійної класифікації (тобто завдання поділу генеральної сукупності на кілька взаємовиключних класів), то описаний нижче процес виконується один раз, одночасно для $N+1$ НВ. Якщо стоїть завдання одержання «нечіткої» класифікації, що припускає можливість приналежності об'єкта відразу декільком класам, і полягає в обчисленні деякій кількісній мірі приналежності об'єкта до класів K_1, \dots, K_N (інакше – задача ранжирування), то процес виконується N разів, для кожної пари НВ.

Серед ознак, що входять до ПН, виконується пошук ДН. Він повинен забезпечувати не тільки прийнятну ФІ, але й ГЗ одержуваних рішень. Для цього знаходиться система ознак, при якій досягається мінімум функціонала $f(\text{ЙПК}, Q2)$, де ЙПК – ймовірність помилки класифікації при використанні деякого набору ознак і обраного метода побудови ВП класифікації; $Q2$ – параметр, що характеризує ГЗ системи.

Для визначення ЙПК використовується будь-який з розроблених критеріїв оцінки байесовської ЙПК (наприклад, на основі відстані Фішера, Бхаттачарія, Колмогорова, дивергенції Кульбака або евристичних показників [6] та ін.).

$Q2$ обчислюється в такий спосіб [7]. Нехай даний деякий фіксований набір з P сіткових ознак $X^{(P)}$ і ЦГМ1. Коміркам ЦГМ1 приписані вірогідно відомі значення деякого геологічного атрибуту (вік порід, назва літологічної відмінності або інше). Будемо умовно розглядати кожен сукупність комірок ЦГМ1 з однаковим значенням атрибуту, як окремий кластер $C_i, i=1..N$, де N – кількість унікальних значень атрибуту. Нехай приналежність об'єктів генеральної сукупності кластерам задається значеннями відповідних комірок ЦГМ1. У цьому випадку $Q2$ може бути розрахований аналогічно $Q1$, як $Q2 = D_{вк}/D_{мк}$, але не для об'єктів НВ, а для всіх об'єктів-комірок сітки.

$Q2$ є мірою компактності розташування об'єктів з однаковим значенням атрибуту в багатомірному ОП і, фактично, характеризує здатність набору ознак відбивати найбільш загальні риси геологічної будови території. Чим нижче значення $Q2$, тим більше геологічно змістовним є набір ознак.

Побудова ДН виконується одним з відомих методів (послідовного приєднання, послідовного відкидання, випадкового пошуку, випадковий пошук з адаптацією, приєднання з відкиданням, «гілок і границь» і т.п.); на кожній ітерації вирішується задача багатокритеріальної оптимізації, що полягає в пошуку сукупності ознак, для якої досягається мінімум функціонала $f(ЙПК, Q2)$. Практично задача вирішується, наприклад, за допомогою вибору рішення з множини парето-оптимальних (таких, що їх не домінують) альтернатив.

Фактично, оптимізація *ЙПК* забезпечує відповідність результатів класифікації наявним відомостям про цільові об'єкти, $Q2$ – геологічним фактам про геологічне середовище в цілому. Знайдений ДН дозволяє досягти прийнятної точки зору ФТ поділу об'єктів різних класів у багатомірному просторі при збереженні його ГЗ (відповідності ЦГМ1).

Далі у просторі знайденого ДН виконується аналіз і корекція відповідних НВ; основна увага на цьому етапі приділяється виявленню й відкиданню (або переміщенню до КВ) об'єктів, що попадають у області, зайняті переважно об'єктами інших класів.

Весь процес пошуку ДН виконується ітераційно, доки на черговій ітерації структура й склад НВ не перестане змінюватися, що свідче про максимально можливе узгодження ДН і НВ, що, в свою чергу, є запорукою отримання адекватних прогнозів.

Прогнозна модель використовується для отримання прогнозів і включає: а) НВ класів і відповідні їм ДН; ВП класифікації; критерії оцінки ФТ, ГЗ і надійності прогнозів, КВ та ЦГМ2.

Найбільш універсальним методом прогнозування в геології є класифікація з навчанням (інакше – контрольована класифікація або із вчителем), коли ВП формуються на підставі аналізу вже класифікованих об'єктів, згрупованих у НВ. Це дозволяє встановити статистичні, логічні або детерміновані зв'язки між геологічним середовищем і даними, що його описують, і використовувати їх

для класифікації. Формування НВ і ДН у рамках побудови ОМ1 і ОМ2 по своїй суті представляє процес пошуку цих зв'язків.

Існує дві основні постановки задачі класифікації з навчанням:

- обчислити кількісну міру приналежності об'єкта до одному з декількох класів, що дозволяє здійснити т.зв. *нечіткий* прогноз.

- визначити клас, для якого міра приналежності об'єкту, що класифікується, найвища; така постановка задачі дозволяє одержати *категорійний* прогноз.

Задача в першій постановці інакше зветься *ранжируванням*, у другий – *розпізнаванням*.

Також ВП можуть створюватися без використання еталонів, на основі виявлених просторових або інших закономірностей, що виявляються методами секвенційного аналізу або пошуку асоціацій.

ОМЗ використовується для виявлення різноманітних зворотних зв'язків та закономірностей між прогнозами, ознаками та цільовими об'єктами та їх інтерпретації з метою формування науково-обґрунтованих рекомендацій для прийняття управлінських рішень щодо подальших геологічних досліджень або гірничих робіт.

Прогнозування локалізації геологічних об'єктів на основі ІПМ.

Прогнозування можна трактувати як процес інтелектуального аналізу геоданих, що включає:

- а) пошук серед значень ознак стійких, геологічно змістовних зв'язків, властивих тільки об'єктам конкретного класу й не характерних для об'єктів інших;

- б) перетворення знайдених зв'язків у форму ВП класифікації;

- в) застосування вирішальних правил для оцінки значень прогнозованого параметра або приналежності об'єктів до класів.

Для кожного набору даних може бути виділена безліч зв'язків, що визначає можливість одержання безлічі альтернативних прогнозів.

Практично будь-яка виявлена в даних закономірність може бути основою для прогнозу. У багатьох ситуаціях справедливо й зворотне – прогноз може бути використаний для пошуку нових закономірностей.

Прогноз, у загальному значенні, є результатом визначення невідомої характеристики (інакше називаної *залежною змінною*) об'єкта (або явища, процесу, події) на основі сукупності відомих характеристик (*незалежних змінних*). Прогнозування виконується у два етапи: на 1-му, використовуючи набори даних з відомими значеннями залежної змінної або виявлені раніше закономірності, будується прогнозна модель; на 2-му вона використовується для передбачення невідомих значень залежної змінної.

Виділення різноманітних зв'язків у більших об'ємах геоданих виконується за допомогою наведених нижче методів інтелектуального аналізу геоданих.

Секвенційний аналіз. Методи секвенційного аналізу є частиною апарата Data Mining і були розроблені, у першу чергу, для цілей маркетингових досліджень. Основна мета їхнього застосування – виявлення стійких послідовностей виникнення деяких подій (у термінах Data Mining – транзакцій), що утворюють ланцюжки. На основі останніх формулюються закономірності, що мають імовірнісний характер і дозволяють передбачати виникнення подій у майбутньому.

В контексті просторових прогнозно-пошукових геологічних задач закономірності, що виявляються методами секвенційного аналізу, можуть формулюватися в такий спосіб:

- на відстані $d_1..d_2$ від об'єкта X в m випадках з M перебуває об'єкт Y (наприклад, в 5-ти випадках з 10-ти на відстані 300–500 м від вузла перетинання розламів перебуває золоторудне тіло);

- однотипні об'єкти X_1, X_2, \dots, X_L в m випадках з M перебувають на відстані $d_1..d_2$ друг від друга (наприклад, кварцові жили північно-східного азимута простягання в 8 випадках з 12 знаходяться на відстані 150–200 м одна від одної).

Можливе знаходження й більше складних просторових закономірностей, що враховують ієрархічну структуру геологічної будови земної кори (наприклад, виявлення періодичності чередування односпрямованих лінійних структур різних рангів).

Частковим випадком секвенційного аналізу є **аналіз асоціацій** (інакше – пошук асоціацій), заснований на виявленні стійких пар або груп одночасно виникаючих подій, або, у контексті прогнозно-пошукових робіт, просторово співпадаючих структур та об'єктів. Фактично, завдання пошуку асоціацій зводиться до завдання секвенційного аналізу при $d_1=d_2=0$. Технологія використовує ряд алгоритмів, заснованих на переборі, що дозволяють знаходити асоціації між значеннями пар або систем ознак, а також між значеннями ознак і геологічних об'єктів. Крім того, що виявлені в результаті застосування апаратів секвенційного аналізу й пошуку асоціацій закономірності можуть мати самостійне наукове й практичне значення, можливо їхнє використання як ВП для виконання прогнозу.

Кластеризація заснована на виявленні певної спільності властивостей у різних об'єктів, що дозволяє групувати об'єкти в деякі підмножини – кластери – в умовах відсутності навчальних вибірок. У рамках представленої технології процедури автоматичної кластеризації використовують як у якості допоміжних (для поділу НВ на підкласи й уточнення їхньої структури), так і при вирішенні інших завдань, наприклад, районування досліджуваної території за комплексом геоданих.

Класифікація з навчанням є найбільш універсальним методом автоматизованого прогнозування. Виконується на основі ПМ, та дозволяє створити цифрові сіткові карти, що відбивають приналежність об'єктів генеральної сукупності (комірок сітки) класам або ступінь приналежності об'єктів одному із класів. В геологічній практиці саме другий підхід представляється більш гнучким та ефективним; для його реалізації будуються ДН для пар НВ класів K_i-K' , $i=1..N$ як описано вище.

Можливість застосування різних вирішальних правил порушує питання вибору найкращого або узгодження (синтезу) декількох прогнозів для одержання єдиного рішення шляхом зведення ансамблю альтернативних прогнозів до одного, узагальненого, при мінімальній втраті інформації. Дане завдання може бути вирішене шляхом ортогоналізації ансамблю з використанням методу головних компонентів. В якості результуючого прогнозу виступає перша головна компонента, що максимізує вибірккову дисперсію уздовж першої координати. Внесок у дисперсію першої компоненти є також однією з мір надійності, усталеності прогнозу.

Висновки. Технологія прогнозування, що базується на побудові та використанні ІПМ, була випробувана на низці рудних та нафтогазових родовищ, що знаходяться у різних геологічних умовах. Одержані результати дали змогу сформулювати наступні висновки:

▪ запропоновані методи побудови ПН та ДН дозволяють знизити неузгодженість між окремими елементами ІПМ, що, як свідчать обчислювальні експерименти, дозволяє знизити помилку класифікації на 5–25%.

▪ запропонована технологія прогнозування дозволяє більш повно використовувати наявну геологічну інформацію, забезпечуючи її інтеграцію до моделей та методів Data Mining;

▪ використання оцінок геологічної змістовності окремих ознак та їх груп є особливо корисним в умовах малих обсягів НВ, дозволяючи знизити ризик одержання геологічно беззмістовних (але формально точних) прогнозів;

▪ формування й використання інтегральних прогнозних моделей (ІПМ), що засновані на описових і прогнозних моделях Data Mining, та поєднують математичні, геоіконічні і геологічні моделі, різномірні дистанційні геодані і геологічні факти, та критерії їхньої взаємної узгодженості, дозволяє отримувати більш вірогідні та адекватні прогнози.

Список літератури

1. Пивняк Г.Г. ГИС-технология интегрированного анализа разнородных и разноуровневых геоданных / Г.Г.Пивняк, Б.С.Бусыгин, С.Л. Никулин // Доп. НАН України. – 2007. – №6. – С.121–128.
2. Kantardzic M. Data Mining: Concepts, Models, Methods, and Algorithms / M.Kantardzic. – John Wiley & Sons. – 2003. – 343 P.
3. Матвейкин В.Г. Информационные системы интеллектуального анализа / В.Г.Матвейкин, Б.С.Дмитриевский, Н.Р.Ляпин – М.: Машиностроение, 2008. – 92 с.
4. Бусыгин Б.С. Формирование интегральных прогнозных моделей при решении геологически задач / Б.С.Бусыгин, С.Л.Никулин // Матеріали міжнародної науково-технічної конференції ІТММ-2013. – Дніпропетровськ. – С. 77–79.
5. Canny J. A computational approach to edge detection/ Canny J. // IEEE Trans. PAMI. – 1986. – V. 8. – P. 34–43.
6. Бусыгин Б.С. Распознавание образов при геолого-геофизическом прогнозировании. / Б.С.Бусыгин, Л.В.Мирошниченко. – Днепропетровск: Изд-во ДГУ. –1991. –168 с.
7. Нікулін С.Л. Використання критерію геологічної змістовності при побудові інформативних систем ознак в задачах геологічного прогнозу / С.Л. Нікулін, Г.С. Підгорна // Тези доповідей СНКПМІ-2013. – Львів. – 2013. с 63–64.

*Рекомендовано до публікації д.геол.н. Приходченком В.Ф.
Надійшла до редакції 25.02.2013*